Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structure of $N^5$-carboxyaminoimidazole ribonucleotide synthase (PurK) from *Bacillus anthracis*

**Micheal L. Tuntland,[a] Michael E. Johnson,[b] L. W.-M. Fung[a]* and Bernard D. Santarsiero[b]***

[a]Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA, and [b]Center for Pharmaceutical Biotechnology, University of Illinois at Chicago, Chicago, IL 60607, USA

Correspondence e-mail: lfung@uic.edu, bds@uic.edu

Received 19 May 2011
Accepted 19 July 2011

**PDB Reference:** $N^5$-carboxyaminoimidazole ribonucleotide synthase, 3q2o.

The apo structure of $N^5$-carboxyaminoimidazole ribonucleotide synthase (PurK) from *Bacillus anthracis* (*ba*PurK) with $Mg^{2+}$ in the active site is reported at 1.96 Å resolution. PurK is an enzyme in the purine-biosynthetic pathway, unique to prokaryotes, that converts 5-aminoimidazole ribonucleotide to $N^5$-carboxyaminoimidazole ribonucleotide and has been suggested as a potential antimicrobial drug target. Two interesting features of *ba*PurK are a flexible B-loop (residues 149/150–157) that is in close contact with the active site and the binding of $Mg^{2+}$ to the active site without additional ligands.

## 1. Introduction

$N^5$-Carboxyaminoimidazole ribonucleotide synthase (PurK) is the enzyme in the purine-biosynthetic pathway that converts 5-aminoimidazole ribonucleotide (AIR) to $N^5$-carboxyaminoimidazole ribonucleotide ($N^5$-CAIR) (Zhang *et al.*, 2008). It represents a divergent step in which prokaryotes and eukaryotes differ; eukaryotes use AIR carboxylase (PurE class II). It has been shown that *Escherichia coli* becomes auxotrophic when PurK is removed (Thoden *et al.*, 2008). Thus, PurK has been suggested to be a potential target for antimicrobial drug discovery (Firestine *et al.*, 2009). The enzyme, which is a member of the superfamily of ATP-grasp proteins (Thoden *et al.*, 1999; Li *et al.*, 2009), has a conserved and highly flexible B-loop near its active site (Thoden *et al.*, 2008). Structural studies of PurK have provided insight into the mechanism and the rearrangement of the active site. However, the structure of the enzyme with only $Mg^{2+}$, but without ligands such as AIR, adenosine triphosphate (ATP) and/or bicarbonate, in the active site is not available for PurK from any bacterial species. Here, we report the structure of PurK from *Bacillus anthracis* (a Gram-positive organism that causes the disease anthrax and is classified as a category A biological agent) with only $Mg^{2+}$ in the active site to 1.96 Å resolution.

## 2. Methods

### 2.1. Protein expression, purification and analysis

The *B. anthracis* PurK (*ba*PurK) gene (383 amino-acid residues) was cloned into a pDEST15 expression vector using Gateway technology (Invitrogen). A thrombin cleavage site was introduced between the GST tag and PurK to allow the removal of GST from the expressed fusion protein. The DNA sequence of the gene in the expression vector was verified.

Cell growth was carried out in a fermenter (BioFlo 110, New Brunswick Scientific) at 310 K in TB medium (2 l) with ampicillin (0.1 g ml$^{-1}$) until the $OD_{600}$ reached 1.1 and was then induced with IPTG (0.25 g or 0.5 m*M*) followed by additional growth for 2 h. The cells were harvested and a suspension of the cells in Triton X-100 buffer was sonicated to lyse the cells. The fusion protein was purified from the lysate by affinity column (GSH resin, Sigma–Aldrich) chromatography. The GST-fusion protein was incubated at 310 K for 1 h with bovine thrombin (BioPharm Laboratories) and GST was removed by the affinity column to give *ba*PurK. The *ba*PurK sample was dialyzed in Tris buffer (10 m*M* Tris pH 8.0 with 200 m*M* NaCl), concentrated to 260 µ*M* and further purified, as well as analyzed, using a Superdex 200 (GE Healthcare) column at 277 K. The column was calibrated with blue dextran (2000 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), RNAse A (13.7 kDa) and aprotinin (6.5 kDa) (all from GE Healthcare) on an FPLC system (ÄKTA, GE Healthcare). The purity of the prepared protein was determined by SDS–PAGE. The mass was determined by high-resolution/high-mass mass-spectrometric methods (LTQ-FT spectrometer, Research Resources Center, University of Illinois).

### 2.2. Crystallization

Crystallization was carried out using the hanging-drop vapor-diffusion method. The protein (725 µ*M*) was in Tris buffer. The reservoir solution was an optimized crystallization solution consisting of 0.1 *M* MES buffer pH 6.5 containing 9% glycerol, 10% PEG 6000 and 5 m*M* DTT. Silver Bullets Bio additive No. 17 (Hampton Research; 0.2% nicotinic acid, 0.2% inosine 5′-monophosphate disodium salt, 0.2% gibberellin $A_3$, 0.2% *O*-phospho-L-tyrosine and 0.2% caffeine in 20 m*M* HEPES pH 6.8) was added, with volume proportions of 1.0, 0.5 and 0.5 µl protein, reservoir and additive solutions, respectively. Crystals formed and appeared as thin plates after 3 d at 293 K. A solution consisting of 6 m*M* $MgCl_2$ and 3 m*M* ATP/ADP was used for soaking. Apo crystals were soaked for 15 s using a 1:1 ratio of soaking:reservoir solution.

### 2.3. Data collection and structure determination

Data were collected at SER-CAT 22-ID (Southeastern Regional Collaborative Access Team, Advanced Photon Source, Argonne National Laboratory) by flash-cooling the crystals in liquid nitrogen. No additional cryoprotectant was added as the crystallization conditions contained both PEG and glycerol at sufficiently high concentrations. A 300 mm MAR CCD detector was used for data collection and *XDS* (Kabsch, 2010) was used for indexing, integration and scaling. *Phaser* (McCoy *et al.*, 2007) was used for molecular replacement using a homology model built from PurK from *Aspergillus clavatus* (PDB entry 3k5h; Thoden *et al.*, 2010) and resulted in a model with two independent chains. *Coot* (Emsley & Cowtan, 2004) was used for manual rebuilding and *REFMAC* (Murshudov *et al.*, 2011) was used for refinement in *CCP*4 (Winn *et al.*, 2011). Details of the data collection and structure solution are given in Table 1. *MolProbity* (Chen *et al.*, 2010) was used for structure validation. Analysis of the secondary structure and dimer interface was performed using *PDBsum* (Laskowski, 2009). Structural representations were generated using *PyMOL* (v.2r3pre; Schrödinger LLC). Sequence alignment was carried out using *FASTA* (Pearson & Lipman, 1988) and sequence comparison was carried out using *Jalview* v.2.6.1 (Waterhouse *et al.*, 2009) and *ClustalW* (Larkin *et al.*, 2007). The sequence was aligned with that of *E. coli* PurK to determine the boundaries of the domains defined in ATP-grasp family members (Li *et al.*, 2009) and was aligned and compared with those of *E. coli* and *A. clavatus* to determine the location of the B-loop. The structure has been deposited in the Protein Data Bank (PDB entry 3q2o).

**Table 1**
Crystal data for *ba*PurK (PDB entry 3q2o).

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Data collection | |
| Space group | $P2_12_12_1$ [No. 19] |
| Unit-cell parameters (Å) | $a$ = 57.43, $b$ = 82.45, $c$ = 166.62 |
| Resolution range (Å) | 19.96–1.96 (2.07–1.96) |
| No. of unique reflections | 56242 |
| Completeness (%) | 96.9 (81.3) |
| Multiplicity | 7.6 (4.9) |
| $\langle I/\sigma(I)\rangle$ | 14.7 (2.7) |
| $R_{\rm merge}$ (%) | 10.6 (75.0) |
| Wilson $B$ factor (Å$^2$) | 34.2 |
| Refinement | |
| No. of protein residues | 754 |
| No. of water molecules | 379 |
| No. of magnesium ions | 2 |
| $R_{\rm work}/R_{\rm test}$ | 19.5/23.5 |
| R.m.s.d. from ideal values | |
| Bond lengths (Å) | 0.009 |
| Bond angles (°) | 1.11 |
| Torsion angles (°) | 5.34 |
| Chiral centers (Å$^3$) | 0.075 |
| Planes (Å) | 0.004 |
| Average $B$ factor (Å$^2$) | 28.5 |
| Ramachandran plot† | |
| Favorable | 731 [97.5%] |
| Acceptable | 18 [2.4%] |
| Outliers | 1 [0.1%] |

† The number of residues is listed, with the percentage given in square brackets.

## 3. Results

The purity of the recombinant *ba*PurK sample was >95% and its mass was 43 157.7 Da (the expected mass was 43 157.4 Da). The hydrodynamic mass from gel filtration was 86.3 kDa. Thus, PurK is a dimer in solution.

*ba*PurK–$Mg^{2+}$ crystallized in space group $P2_12_12_1$ as a dimer with two independent chains (*A* and *B*) in the asymmetric unit and the structure was refined to 1.96 Å resolution (Table 1), with the last six residues not observed in the density map. The topology of each chain is globular, with 16 helices and 17 $\beta$-strands (forming three $\beta$-sheets). The three ATP-grasp domains in each monomer (chain *A* or *B*) are the A-domain (residues 1–120), B-domain (residues 121–186) and C-domain (residues 187–383), with a nine-residue B-loop (residues

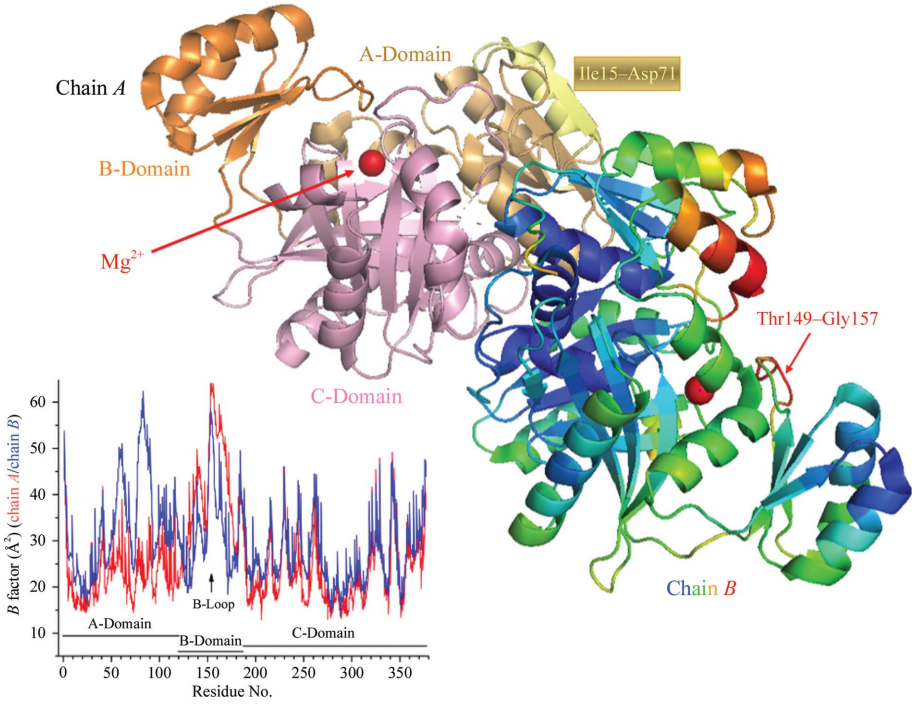


**Figure 1**
*ba*PurK as crystallized in dimer form. Chain *A* is shown with the three domains that define an ATP-grasp protein (A-domain, sand; B-domain, orange; C-domain, pink). Residues 51–71, shown as a deletion in *E. coli* through sequence alignment, are colored pale yellow. Chain *B* is essentially identical to chain *A*. Magnesium cations are shown as red spheres. Chain *B* is depicted by a color spectrum ranging from dark blue for low average *B* factors (about 11 Å$^2$ in the C-domain) to dark red for high average *B* factors [about 63 Å$^2$ for the B-loop (residues 149–157) in the B-domain]. *B* factors are plotted by residue for both chains in the inset plot (lower left). The two chains show similar *B* factors except for residues ∼74–92, for which chain *B* shows higher mobility.

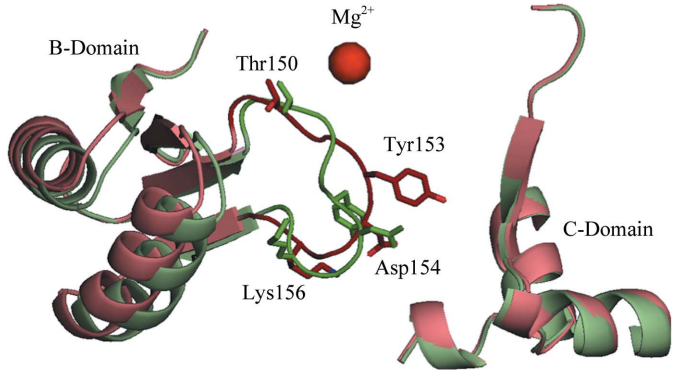


**Figure 2**
Overlay of chain *B* (green) onto chain *A* (red) in the B-loop region (residues 149–157) of the B-domain. The conformations of the B-loop in chain *A* and in chain *B* differ significantly, with residues such as Tyr153, Asp154 and Lys156 of the two chains in different positions. The $\alpha$-helices and $\beta$-strands upstream and downstream of the B-loop in the B-domain also exhibit slightly differing conformations in chains *A* and *B* (left side of the figure). Note that in the same orientation the superposition of the C-domains of the two chains is virtually identical.

149–157; TTGGYDGKG) in the B-domain (Fig. 1). The C$^\alpha$ *B*-factor values range from about 11 to 63 Å$^2$, with the highest values in the B-loop (Fig. 1). The main-chain atoms in the C-domain are the least mobile, while the main-chain atoms in the B-domain are the most mobile.

Chains *A* and *B* are very similar, with a root-mean-square displacement (r.m.s.d.) of 0.4 Å for C$^\alpha$ atoms. The region of least similarity is the flexible B-loop, leading to slightly different positions for the B-domain in chains *A* and *B* and slightly different conformations of some of the side chains in the B-loop, such as Tyr153, Asp154 and Arg156 (Fig. 2). These differences suggest that the B-loop conformation can easily be affected by local environment and interactions. The interface of chains *A* and *B* involves 57 residues, 28 from chain *A* and 29 from chain *B*, and includes 16 hydrogen bonds and 123 nonbonded contacts, with a contact surface area of 3150 Å$^2$.

Each chain has one $Mg^{2+}$ ion coordinated to five O atoms in a distorted trigonal bipyramidal configuration. In chain *A*, $Mg^{2+}$ is coordinated to the four O atoms of four water molecules and the backbone carbonyl O atom of Leu269 in the C-domain (Fig. 3). In chain *B*, $Mg^{2+}$ is coordinated to two O atoms from two water molecules, two backbone carbonyl O atoms from Leu269 and Thr104 and the O$^\varepsilon$ atom of Glu110 in the A-domain. The coordination distances between $Mg^{2+}$ and the O atoms range from 1.94 to 2.47 Å. The $Mg^{2+}$-binding site is in a cavity with several charged residues, including Arg107 at the A-domain C-terminus, Lys147 in the B-domain region upstream of the B-loop, Glu82 and Lys183 in the B-domain C-terminus and Glu190 in the C-domain N-terminus. This cavity is generally considered to be the active site of PurK from *E. coli* (*ec*PurK) and *A. clavatus* (*ac*PurK), in which $Mg^{2+}$ as well as ADP, ATP, P$_i$ *etc.* are found (Thoden *et al.*, 1999, 2008, 2010). The B-loop of *ba*PurK, although not resolved in the published structure of *ec*PurK without ligand (Thoden *et al.*, 1999), is at the opening of the cavity and is likely to play an important role in the ligand binding to *ba*PurK.

## 4. Discussion

We have obtained the structure of *ba*PurK with only $Mg^{2+}$ in the active site at a resolution of 1.96 Å. This is the only currently available structure of PurK with $Mg^{2+}$ but without ligands such as ADP, ATP or AIR in the active site. The ligands in the previously published structures of *ec*PurK and *ac*PurK include ADP, ATP, AIR, P$_i$, Cl$^-$, $SO_4^{2-}$ *etc.* individually or in combinations with each other or with $Mg^{2+}$ (PDB entries 1b6r, 1b6s, 3eth, 3etj, 3k5h and 3k5i; Thoden *et al.*, 1999, 2008, 2010). We suggest that this structure with only $Mg^{2+}$ in the active site is an important starting structure for

understanding the subsequent binding of AIR, $HCO_3^-$ and ATP to the active site. Sequence alignments of *ba*PurK, *ec*PurK and *ac*PurK (431 of 572 residues) show about 30% identity and 60% similarity (Fig. 4). The 63 conserved residues are scattered throughout the sequence, including those in the B-loop. The B-loop is considered to be an important feature of the structure since it is in contact with the phosphate group, but it is highly flexible and generally not observed without ligand, as in the reported X-ray structure of *ec*PurK with $SO_4^{2-}$. In this study, the B-loop of *ba*PurK consists of nine residues (residues 149–157; T**T**GG**YDG**K**G**), with the residues shown in bold being conserved and exhibiting high *B* factors (above 60 Å$^2$). In *ec*PurK and *ac*PurK structures with ligands where the B-loop is observed in the structure, the B-loop appears to be less flexible (Thoden *et al.*, 1999, 2008, 2010). This B-loop may be selected with varying flexibility to allow ligands to enter the adjacent active site, with the loop then becoming less flexible and more ordered to facilitate the conversion of AIR and bicarbonate in the presence of ATP to the product $N^5$-CAIR.

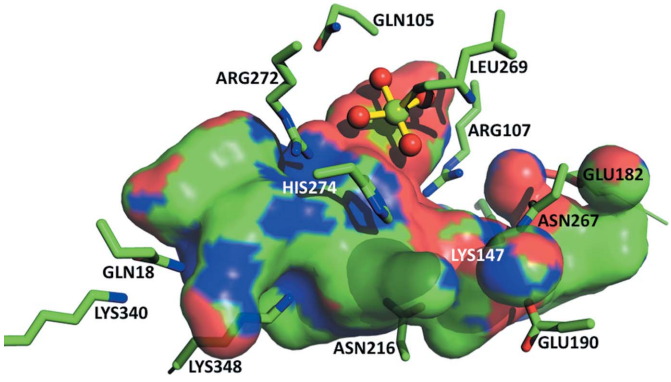


**Figure 3**
In chain *A*, $Mg^{2+}$ is coordinated to the four O atoms of four water molecules and the backbone carbonyl O atom of Leu269 in the C-domain. The coordination distances between $Mg^{2+}$ and the O atoms range between 1.94 and 2.47 Å. The $Mg^{2+}$-binding site is on the rim of a large cavity, depicted here by a bound surface with basic (blue), acidic (red) and hydrophobic (green) residues, and surrounded by several charged residues, including Arg107 at the A-domain C-terminus, Lys147 in the B-domain region upstream of the B-loop, Glu182 in the B-domain C-terminus and Glu190 in the C-domain N-terminus. The surface is drawn with 20% transparency to reveal residues behind the cavity surface (Lys147, Glu182 and Asn216).

A relatively long segment, residues 51–71, in *ba*PurK (Fig. 1) is missing in *ec*PurK and exhibits little homology with that of *ac*PurK (Fig. 4). This *ba*PurK segment in the A-domain near the N-terminus is comprised of a $\beta$-strand followed by a helix and is flanked by a $\beta$-strand–helix–$\beta$-strand upstream and a $\beta$-strand–helix downstream ($\underline{\beta\alpha\beta}$u$\underline{\boldsymbol{\beta\alpha\beta}}\underline{\alpha}$, where the underlined regions are similar in all three PurKs, the region in bold is missing in *ec*PurK and 'u' represents an unstructured segment; Figs. 1 and 4). The corresponding secondary-structural elements are $\underline{\beta\alpha\beta}\underline{\alpha\alpha}\underline{\beta\alpha}$ in *ec*PurK (Thoden *et al.*, 2008) and $\underline{\beta\alpha\beta}\underline{\alpha}\underline{\boldsymbol{\beta\alpha\beta}}\underline{\alpha}$ in *ac*PurK (Thoden *et al.*, 2010). Since there are currently no structures of *ec*PurK or *ac*PurK with only $Mg^{2+}$ bound to the active site, it is not clear whether the deletion of this 20-residue segment causes unique conformational differences that lead to significant functional differences in *ec*PurK amongst the three species. In the following segment (Glu76–Leu94) the *B* factors vary considerably between chain *A* and chain *B*. Values of 18–30 Å$^2$ are observed for chain *A*, while values of 38–61 Å$^2$ are seen for chain *B* (inset in Fig. 1). This is a consequence of crystal packing, with this segment in chain *A* having a compact association with a symmetry-related chain *A*, but with that in chain *B* having no close symmetry-related protein contacts. Additionally, this segment could significantly influence enzyme activity and mechanism, since it contains three conserved residues (Glu76, Glu78 and Leu85; Fig. 4). Interestingly, there is a 40-residue insertion in *ac*PurK at the C-terminal end (Fig. 4). The significance or lack of significance of these differences awaits future examination.

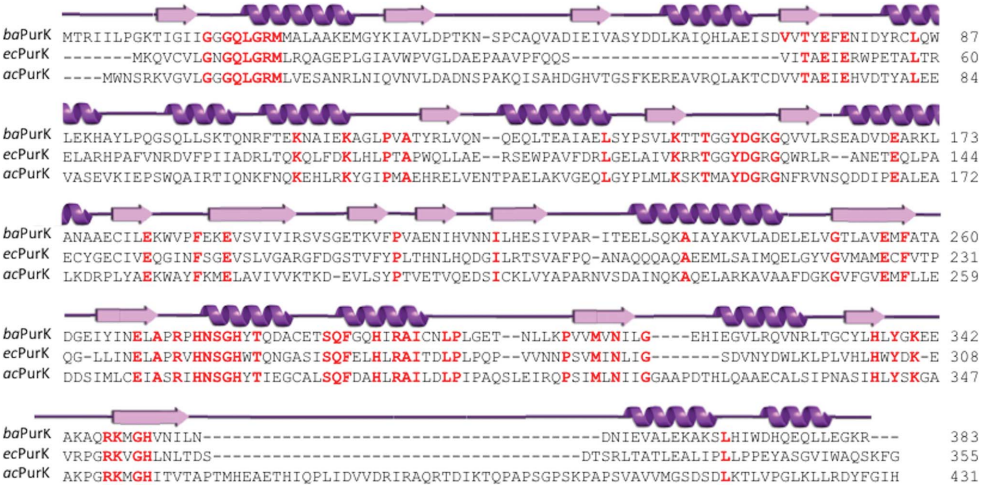


**Figure 4**
Sequence alignment and secondary-structure assignment of *ba*PurK (accession No. YP002864), *ac*PurK (accession No. EAW10687) and *ec*PurK (accession No. P09029) showing the conserved residues (red), the *ec*PurK deletion and the *ac*PurK insertion segment. The sequences of *ba*PurK and *ec*PurK show 32.7% identity and 61.9% similarity, while those of *ba*PurK and *ac*PurK show 33.2% identity and 67.0% similarity. The sequence of *ac*PurK (572 residues) is considerably longer then those of *ba*PurK and *ec*PurK.

The $Mg^{2+}$ coordination sphere for *ec*PurK with $Mg^{2+}$ and ADP (PDB entry 1b6s) or ATP (PDB entry 3eth) includes two O atoms from the ADP or ATP phosphates, while the remaining three to four O atoms are from the $O^{\varepsilon}$ atoms of Glu226 and Glu238 (Thoden *et al.*, 1999, 2008). These two glutamic acid residues are conserved in all three species and correspond to Glu255 and Glu268 in the active site of *ba*PurK (Fig. 3). Thus, we suggest that the pre-coordination of $Mg^{2+}$ to the O atoms of water molecules, backbone carbonyl O atom(s) (of Leu269 in chain *A* and of Leu269 and Thr104 in chain *B*) and the $O^{\varepsilon}$ atom of Glu110 (in chain *B*)

in the absence of ligand(s) would facilitate the eventual binding of ATP, bicarbonate and AIR to the active site and assume a final coordination with the O atoms of Glu255 and Glu268, a residue next to Leu269, and the O atoms of phosphate in ATP or ADP. Charged or polar residues around Thr104 and Glu110, such as Arg107 or Lys111, may also be involved in active-site ligand binding in *ba*PurK.

## 5. Conclusions

The structure of *ba*PurK with only $Mg^{2+}$ in the active site shows $Mg^{2+}$ coordinated to a set of O atoms different from those in published PurK structures that have ATP or ADP in the active site. Our structure further shows a longer B-loop, suggesting a more flexible B-loop in the absence of ligands in the active site. We propose that $Mg^{2+}$ and the B-loop are positioned uniquely in *ba*PurK to facilitate both the subsequent binding of substrate molecules (AIR and bicarbonate) and cofactor (ATP) and the conversion to the product molecules $N^5$-CAIR, ADP and $P_i$.

This work was supported in part by the Transformational Medical Technologies program contract HDTRA1-11-C-0011 from the Department of Defense Chemical and Biological Defense program through the Defense Threat Reduction Agency (DTRA). Data were collected on the Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38. One of us (BDS) acknowledges support from the UIC Center for Clinical and Translational Science (CCTS), award number UL1RR029879, from the National Center for Research Resources.

## References

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Firestine, S. M., Paritala, H., McDonnell, J. E., Thoden, J. B. & Holden, H. M. (2009). *Bioorg. Med. Chem.* **17**, 3317–3323.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–D359.
Li, H., Fast, W. & Benkovic, S. J. (2009). *Protein Sci.* **18**, 881–892.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Pearson, W. R. & Lipman, D. J. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
Thoden, J. B., Holden, H. M. & Firestine, S. M. (2008). *Biochemistry*, **47**, 13346–13353.
Thoden, J. B., Holden, H. M., Paritala, H. & Firestine, S. M. (2010). *Biochemistry*, **49**, 752–760.
Thoden, J. B., Kappock, T. J., Stubbe, J. & Holden, H. M. (1999). *Biochemistry*, **38**, 15480–15492.
Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zhang, Y., Morar, M. & Ealick, S. E. (2008). *Cell. Mol. Life Sci.* **65**, 3699–3724.